

Frank Seifart\*, Jan Strunk, Swintha Danielsen, Iren Hartmann, Brigitte Pakendorf, Søren Wichmann, Alena Witzlack-Makarevich, Nikolaus P. Himmelmann and Balthasar Bickel

# The extent and degree of utterance-final word lengthening in spontaneous speech from 10 languages

<https://doi.org/10.1515/lingvan-2019-0063>

Received November 27, 2019; accepted April 2, 2020; published online February 9, 2021

**Abstract:** Words in utterance-final positions are often pronounced more slowly than utterance-medial words, as previous studies on individual languages have shown. This paper provides a systematic cross-linguistic comparison of relative durations of final and penultimate words in utterances in terms of the degree to which such words are lengthened. The study uses time-aligned corpora from 10 genealogically, areally, and culturally diverse languages, including eight small, under-resourced, and mostly endangered languages, as well as English and Dutch. Clear effects of lengthening words at the end of utterances are found in all 10 languages, but the degrees of lengthening vary. Languages also differ in the relative durations of words that precede utterance-final words. In languages with on average short words in terms of number of segments, these penultimate words are also lengthened. This suggests that lengthening extends backwards beyond the final word in these languages, but not in languages with on average longer words. Such typological patterns highlight the importance of examining prosodic phenomena in diverse language samples beyond the small set of majority languages most commonly investigated so far.

**Keywords:** final lengthening, word duration, language documentation, prosodic typology

## 1 Introduction

It is well known that articulation tends to slow down towards the end of prosodic units such as intonational phrases and utterances. This is referred to as final lengthening, prepausal lengthening, domain-final lengthening or preboundary lengthening. It is “considered by many to be universal” (Fletcher 2010: 540; see also Lindblom 1968; 1979; Vaissière 1983), but there are also indications that “the degree and extent of lengthening varies among languages” (Fletcher 2010: 540). The current study investigates final lengthening effects on durations of utterance-final words, and words preceding these, expanding earlier corpus-based studies of word durations in English (e.g., Bell et al. 2003; Yuan et al. 2006) to a broad sample of languages.

---

\*Corresponding author: Frank Seifart, Leibniz-Zentrum Allgemeine Sprachwissenschaft, Berlin, Germany; Dynamique Du Langage (CNRS & Université de Lyon), Lyon, France; and University of Cologne, Cologne, Germany, E-mail: seifart@leibniz-zas.de. <https://orcid.org/0000-0001-9909-2088>

Jan Strunk and Nikolaus P. Himmelmann: University of Cologne, Cologne, Germany. <https://orcid.org/0000-0001-8546-1778> (J. Strunk). <https://orcid.org/0000-0002-4385-8395> (N.P. Himmelmann)

Swintha Danielsen: CIHA, Santa Cruz de la Sierra, Bolivia

Iren Hartmann: Leipzig University, Leipzig, Germany

Brigitte Pakendorf: Dynamique Du Langage (CNRS & Université de Lyon), Lyon, France

Søren Wichmann: Leiden University, Leiden, The Netherlands; Kazan Federal University, Kazan, Russia; and Beijing Language University, Beijing, China

Alena Witzlack-Makarevich: Hebrew University of Jerusalem, Jerusalem, Israel. <https://orcid.org/0000-0003-0138-4635>

Balthasar Bickel: University of Zurich, Zurich, Switzerland

From a language production perspective, final lengthening may reflect planning efforts of the following speech constituents (Oller 1979) and dynamic effects on the activation time-course of articulatory gestures (Byrd and Saltzman 2003). Both of these factors would lead one to expect that such lengthening should be observed across different languages. Final lengthening can also be viewed as a listener-oriented strategy to signal different levels of constituency (e.g., Turk and Shattuck-Hufnagel 2000), which allows for cross-linguistic and cross-cultural variation (Ordin et al. 2017).

Final lengthening occurs at the right edge of various types of units in the hierarchy of prosodic phrasing (Shattuck-Hufnagel and Turk 1996), such as prosodic words, (full) intonational phrases, and utterances, and the higher the prosodic domain, the stronger the effects (Michelas and D’Imperio 2010; Wightman et al. 1992). Prosodic phrasing is manifested in phrase type-specific and language-specific combinations of a variety of features. Final lengthening and pauses are prominent features in many languages, but they may combine with, e.g., pitch movements (Jun 2014) and changes in voice quality and/or intensity (Himmelman et al. 2018; Himmelman and Ladd 2008: 252).

How far back from a boundary is the extent – also called “domain” (Cambier-Langeveld 1997; Turk and Shattuck-Hufnagel 2007) or “temporal scope” (Byrd et al. 2006) – of final lengthening? There is general consensus that it is “largely, though not entirely, limited to the boundary-adjacent segments” (Byrd et al. 2006: 1590) like the rhyme of phrase-final syllables, with potentially additional lengthening extending up to penultimate syllables or the final foot (Fletcher 2010: 545). Evidence from Italian, Finnish, and Japanese shows that how far final lengthening extends backwards may depend on language-specific stress and vowel quantity characteristics (Cho 2016: 125). Lengthening throughout phrase-final disyllabic words has been observed in English and Hebrew (Berkovits 1994; Byrd et al. 2006) and up to the initial syllable in the English word *banana* (Cho et al. 2013).

Regarding final-lengthening effects on words as a whole – as investigated in the current study – Bell et al.’s (2003) analysis of function words in spontaneous speech from the Switchboard corpus of American English telephone conversations (Godfrey et al. 1992) clearly shows lengthening of utterance-final words, consistent with results from Yuan et al.’s (2006) analysis of all words in the same corpus. Very few studies have investigated potential final-lengthening effects extending beyond final words. Yuan et al.’s (2006) analyses of Switchboard data strongly suggest that penultimate words in utterances are also affected, but not antepenultimate words and beyond. Analyses of Finnish read speech also found final-lengthening effects in penultimate words (Hakokari et al. 2005) (on the temporal domain of accentual lengthening, see Turk and White 1999).

How much longer are final (portions of) words compared to non-final ones, i.e., what is the degree of final lengthening? Utterance-final syllables in experimental data from English, Spanish, and German have been reported to be up to 75% longer than non-final syllables (Delattre 1966). Finnish phonemes were found to be 23–51% longer (depending on the type of phoneme) in phrase-final words than in phrase-medial words, and 3–6% longer in penultimate words than in medial words (Hakokari et al. 2005). Regarding the lengthening of words as a whole – as investigated in the current study – English function words in final positions in the Switchboard corpus have been reported to be 23% longer (Bell et al. 2003: 1020) after controlling for other factors, including contextual probabilities. Yuan et al.’s (2006) results suggest that, in this corpus, final words are on average lengthened by approximately 50%, with similar results for different parts-of-speech, without, however, including control factors. For penultimate words, Yuan et al.’s (2006) results indicate roughly 10% lengthening.

Systematic cross-linguistic comparative studies on final lengthening are extremely rare and not recent. This is surprising, because an early comparison of English, French, German, and Spanish revealed striking differences: For instance, the increase in duration of final versus non-final stressed open syllables was 74% in English versus only 21% in Spanish (Delattre 1966: 194).

The current study investigates cross-linguistic variation in the extent and degree of final lengthening in a novel data type: corpora of spontaneous speech in 10 genealogically, areally, and typologically diverse languages (Figure 1). Eight of these corpora stem from recent efforts to document under-resourced, small, and often endangered languages in annotated multimedia corpora in terms of Himmelman (1998). All data were transcribed, translated, and morphologically analyzed by language experts, segmented into utterances (by a combination of manual annotation and automatic pause detection), and time-aligned at the word level. Our methodology is tailored to the under-resourced nature of most of the languages studied here, specifically to the

lack of phonemic or phonetic transcription and syllabification. This has two implications for the comparability of our results with previous, mostly experimental studies on well-resourced languages: Firstly, our baseline measures on word length and speech rate are based on orthography as a proxy for phonological segments. This is relatively unproblematic because the orthographies are grounded in careful phonological analyses by language experts in consultation with native speakers. Secondly, we focus on lengthening of (orthographic) words as a whole (following, e.g., Bell et al. 2003; Yuan et al. 2006), not syllables or segments. However, the fact that final syllables and segments within words are disproportionately strongly lengthened is crucial for our comparative analyses, since we therefore expect shorter words to be more strongly affected by final lengthening than longer words. Note also that measuring word durations captures final lengthening irrespective of the precise location of lengthening within final words with, e.g., Bantu languages lengthening penultimate, rather than ultimate syllables (Hyman 2013).

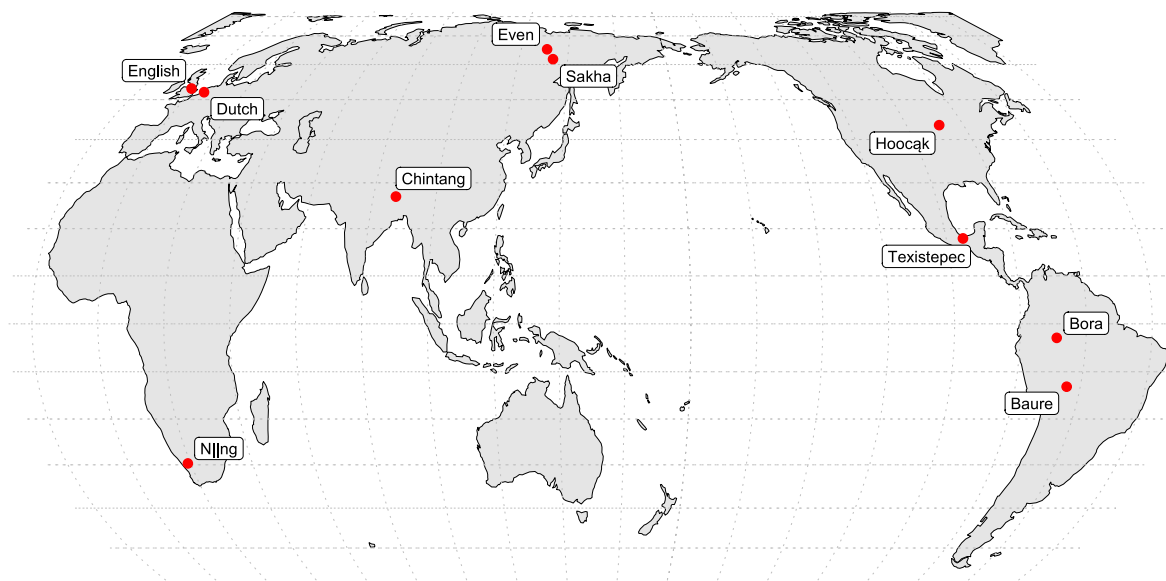
The overarching goal of the current study is to bring spontaneous corpus data from under-resourced languages to bear on our understanding of lengthening at prosodic boundaries. Our first specific aim is to test the widespread assumption that cross-linguistically, there is a consistent lengthening effect on the durations of utterance-final words. Secondly, we investigate whether cross-linguistically lengthening extends to words preceding utterance-final words. Finally, we describe cross-linguistic variation in the degree of word lengthening at utterance boundaries.

## 2 Data and methods

### 2.1 Data

#### 2.1.1 Corpus characteristics

The language sample used here (Figure 1, Table 1) includes data from the Switchboard corpus of English and the Dutch *Corpus Gesproken Nederlands*, allowing for comparison of our results with earlier studies. From these two corpora, we chose sets of recordings of spontaneous speech from single varieties that approximately match the other eight corpora in terms of total number of words. These eight corpora consist of texts recorded



**Figure 1:** Location of languages included in the study on an Equal Earth map projection (Šavrič et al. 2019).

during fieldwork, mostly traditional or personal narratives. All data are spontaneously spoken, not read or memorized, even if texts stem from local oral traditions.

All data have been orthographically transcribed, translated into a major language and morphologically analyzed and part-of-speech tagged by experts on these languages. Word boundaries as given in the transcriptions follow orthographic conventions, according to which clitics are typically written as affixes, i.e. word units approximate prosodic words. While there may be differences across corpora in the treatment of clitics as affixes or as separate words, the same word segmentation was used for calculating word durations and word length within each language.

Transcriptions in the eight language documentation corpora were time-aligned by language experts at the level of annotation units using ELAN (ELAN developers 2019), mostly for practical purposes, e.g., to display translations as subtitles. For some languages, annotation units usually comprise one clause (e.g., Texistepec, average length: 3.9 words). For others, such units might be better characterized as paragraphs (Hoocąk, 7.9 words on average). These data were further automatically time aligned using the WebMAUS software (Kisler et al. 2012). We did not carry out any training (for which our corpora are too small) or adaptation of WebMAUS to specific languages, but instead used the superset of acoustic models of all languages that MAUS currently supports. This procedure results in fairly accurate word start and end times (for details, see Strunk et al. 2014), but phoneme-level alignments are overall not reliable enough for our purposes. All word start and end times were subsequently manually checked and, where necessary, corrected, with concurrent manual annotation of filled pauses. Corrections of phoneme times could not be carried out because of limited resources, and phoneme times were thus not used. For Dutch and English, we used the time-alignment provided by the creators of these corpora.

### 2.1.2 Data preparation

The entire dataset (see Table 1) was used to calculate word frequencies as control factors for the analyses. To study word durations, we focus on the final four words of utterances only. Since the annotation units in our corpora do not always match utterances, as prosodically defined, we use pauses as an additional criterion to identify utterance boundaries. Specifically, for word durations, we only include four consecutive words that fulfill the following criteria (see Appendix A in the Supplemental Materials for the amount of data retained after each step):

- (i) They are the final four words of annotation units that are at least five words long and these four words are not preceded by a pause of 0.2 s or more. By this criterion, we exclude words that are in utterance-initial position and we exclude lengthening in the vicinity of utterance-internal hesitation pauses.
- (ii) They are the final four words of annotation units that are immediately followed by a pause of at least 0.2 s. This criterion ensures that the utterance boundaries used in the analysis in fact correspond to prosodic boundaries, rather than being inserted by the annotator for semantic, syntactic, or practical reasons only.
- (iii) They do not include any silent pauses longer than 0.2 s, any filled pauses, any disfluent words, any false starts, or any word that could not be identified during transcription. This criterion excludes utterances shorter than four words and words that are either disfluencies themselves or affected in their duration by disfluencies in their vicinity.

Through this procedure, we retain only utterance-final words, penultimate words, and two words considered as medial (Figure 2), while making sure that the duration of these words is not affected by phrase-initial lengthening (Keating et al. 2003) or lengthening in the vicinity of pauses or disfluencies (Bell et al. 2003). Note however, that our method does not pick up prosodic boundaries within such four-word chunks that are not marked by pauses longer than 0.2 s, but instead by, e.g., pitch or shorter pauses, even though such boundaries may contribute to (utterance-medial) phrase-final lengthening, as discussed in Section 4.

For Texistepec, we applied an additional data selection procedure to account for the use of the quotative verb *d'im* 'he/she/it said', which marks the end of direct speech and occurs 1830 times in the corpus, of which 1219 times in utterance-final position. In these cases, we excluded the quotative verbs since the utterance

**Table 1:** Basic information on corpora and languages, with Glottolog language identification codes (Hammarström et al. 2018). Note that the language we call “Texistepec” here is often also referred to as “Texistepec Popoluca”.

Language	Typology				Corpus					
	Language	Glottocode	Family	Word order	Stress versus tone	Vowel length	Av. Segments/ word	Av. Morphemes/ word	Words (total)	Words (duration study)
Baure	baur1253	Arawakan	VSO	Stress	No	1.86	5.73	17 563	2 992	Danielsen et al. (2009)
Bora	bora1263	Boran	SOV	Tone	Yes	2.21	7.13	29 795	7 080	Seifart (2009)
Chintang	chhi1245	Sino-Tibetan	SOV	Stress	No	1.81	5.14	37 731	5 096	Bickel et al. (2011)
Dutch	dutc1256	Indo-European	SOV	Stress	Yes	n/a	3.85	39 448	8 128	CGN-consortium (2003)
English	stan1293	Indo-European	SVO	Stress	(Tenseness)	1.09	3.70	56 136	8 544	Godfrey et al. (1992)
Even	lamu1253	Tungusic	SOV	Stress	Yes	1.91	5.79	37 394	12 116	Pakendorf et al. (2010)
Hooçak	hoch1243	Siouan	SOV	Stress	Yes	1.71	6.64	23 176	7 440	Hartmann (2013)
Nlɨŋ	nuuu1241	!Uj-Taa	SVO	Stress tone	No	1.14	3.45	25 850	5 112	Güldemann et al. (2011)
Sakha	yaku1245	Turkic	SOV	Stress	Yes	1.68	5.77	31 139	8 560	Pakendorf (2007)
Texistepec	texi1237	Mixe-Zoquean	SOV	Stress	Yes	1.81	5.14	21 315	4 044	Wichmann (1996)

consisting of quoted speech arguably ends before the final quotative verb (for results from the complete Texistepec dataset see Appendixes A–C in the Supplemental Materials).

## 2.2 Statistical modeling

We built multivariate linear mixed-effects models using the R library *lme4* (Bates et al. 2015; R Core Team 2018) to analyze the effect of word position within utterances on word duration, while controlling for other known relevant variables. For optimal comparability of the strength of the fixed effects across individual languages, we carried out parallel analyses applying the same model structure, as given in (1), to our 10 individual corpora (with the only exception that “morphological complexity” was excluded for Dutch, see Section 2.4.4). To keep the model structure constant, we refrained from performing model search, selection of variables, or inclusion of interactions between variables for individual languages. We use an intercept-only random effects structure, because for individual languages, models with more complex random effects structures often did not converge, probably because of data scarcity, which again would make comparisons between languages less reliable and less easy to interpret.

$$(1) \log(\text{word duration}) \sim \text{position} + \log(\text{relative frequency}) + \text{word length} + \text{number of morphemes} + \text{word class} + \log(\text{local speech rate}) + (1|\text{speaker}) + (1|\text{text}) + (1|\text{word type})$$

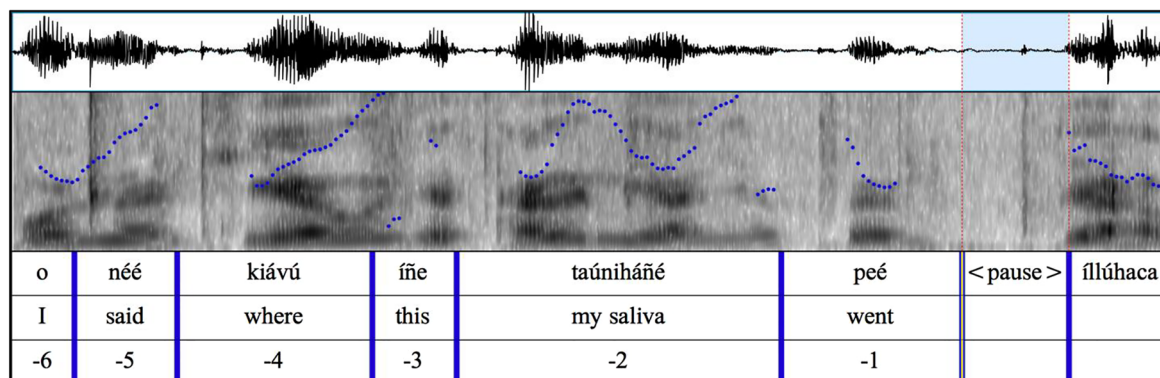
## 2.3 Dependent variable: Word duration

We use word duration (based on manually corrected word start and end times, see Section 2.1.1) as the main dependent variable, following recent work on final lengthening and work on word lengthening/contraction as a function of, e.g., frequency (e.g., Bell et al. 2009; Seyfarth 2014; Sóskuthy and Hay 2017; Tang and Bennett 2018). To capture relative changes in duration rather than absolute ones, we use the (natural) logarithm of word duration.

## 2.4 Independent variables

### 2.4.1 Word position

Our main independent variable of interest is the position of the target word within an utterance. We encode word position in the form of a discrete factor with four levels (−4, −3, −2, −1), see Figure 2. We compared both the average duration of final words and the average duration of penultimate words to the average duration of



**Figure 2:** Bora utterance boundary set by a human annotator and marked by a pause. Words -4 to -1 are included in the analyses. Original file archived at <https://hdl.handle.net/1839/00-0000-0000-000D-5BF7-6>.



medial words (positions  $-3$  and  $-4$ ) to obtain independent and directly comparable estimates of lengthening for each of them. We additionally compared durations of medial words ( $-3$  vs.  $-4$ ).

#### 2.4.2 Word frequency

The frequency, contextual predictability, and informativity (average contextual predictability) of words are well-known determinants of word duration (Aylett and Turk 2004; Bell et al. 2009; Seyfarth 2014; Tang and Bennett 2018, among others). However, such scores are notoriously difficult to obtain for small corpora. The minimum corpus size for reliable frequency counts has been estimated at 16 million words (Brysbaert and New 2009: 980), although contextual predictability scores have recently been estimated for an under-resourced language based on only 0.7 million words (Tang and Bennett 2018).

For the languages in the current study, textual material beyond the corpora used here is extremely limited, except for English and Dutch and to some extent Chintang and Sakha. To keep results for all languages comparable, we counted word frequencies only in the corpora we have available. We used the entire corpora (see Table 1), which are about 4.5 times bigger than the subsets of data in which durations are measured. To obtain comparable scores across corpora that vary in size, we use (log-transformed) relative frequencies calculated by language/corpus.

The validity of frequency measures based on such small corpora might be questioned. However, the frequency scores we obtained yield the expected significant negative effects on word durations in seven out of the 10 languages, with nonsignificant results in two and an unexpectedly inverse effect in one language (see Appendix C in the Supplemental Materials for results, cf. also Strunk et al. 2020). This overall expected effect indicates that these frequency measures are indeed valid for our type of analysis. We refrain from attempting to calculate (forward or backward) bigram probabilities, as these require even more data than (unigram) frequencies.

#### 2.4.3 Word length

We use word length as a baseline control variable for the duration of words, against which we assess the increase in duration as a function of position in an utterance. In the absence of phonological or phonetic transcriptions and syllabification, word length is measured by the number of orthographic characters as a proxy for the number of phonological segments. For the language documentation corpora, reliance on orthography is justified by close correspondences between phonemic and orthographic representations as devised by language experts. For methodological consistency, we apply the same measure to English and Dutch data, which is further justified by the fact that correlations between word length in orthographic characters and word length in phonological segments are extremely high, even for languages with relatively deep orthographies such as English and Dutch (Piantadosi et al. 2011: 3528). We are aware that our approximation of baseline duration does not distinguish between different kinds of phonological segments or syllable types, but note that our results on English match previous results based on more sophisticated measures closely (see Section 3).

#### 2.4.4 Morphological complexity

The languages in our study display different numbers of morphemes per prosodic word (see Table 1), as annotated by language experts. Inclusion of this factor accounts for earlier findings of lengthening at morpheme boundaries within words (Plag et al. 2017), although it remains unclear how this affects word durations in typologically diverse languages (Strunk et al. 2020; Tang and Bennett 2018). Since the Dutch corpus we used did not include morphological annotation, and providing it would have been beyond the scope of the current project, we excluded this variable from the analysis of Dutch.

### 2.4.5 Word class

Word class is included as a variable here because it has been found that nouns are pronounced more slowly than verbs across languages (Seyfarth 2014: 145–146; Sóskuthy and Hay 2017: 305; Strunk et al. 2020; see also Seifart et al. 2018). We use the word-class category of the lexical root contained in a word (noun vs. verb vs. other), as identified by language-specific criteria. Even though individual words may be nominalized or verbalized, in our data, this occurs in less than 5% of nouns and verbs.

### 2.4.6 Local speech rate

Local speech rate is another control variable for word duration. It is calculated as phonological segments (for which we use orthographic characters as a proxy, see Section 2.4.3) per second, including pauses, in the complete annotation unit surrounding the word whose duration is being modeled (i.e., extending backwards beyond word  $-4$ , see Section 2.1.2). The modeled word's length in phonological segments and its duration in time were excluded from the calculation by subtracting them from the overall length and duration of the annotation unit surrounding it.

## 2.5 Random effects

To account for random variation between speakers and recordings/texts, we included random intercepts for both. We also included per-word random intercepts to model idiosyncrasies of individual word types.

## 3 Results

Results presented in Figure 3 and Table 2 show that in all 10 languages, words in final position have significantly longer durations than medial words. Moreover, lengthening factors and standardized  $\beta$  coefficients are highest for the comparison of final words ( $-1$ ) with medial words ( $-4$  and  $-3$ ). Comparison of penultimate words ( $-2$ ) with medial words yields mixed results across the 10 languages. In four languages – in the two Amazonian languages Baure and Bora, in Chintang, and in Texistepec – there is no statistically significant difference in duration. In three languages – English, Dutch, and in NlIng – penultimate words have significantly longer durations than medial words. In three languages – the two Siberian languages Even and Sakha, and in North American Hoocak – penultimate words have significantly shorter durations than medial words. Moreover, in these three languages words in position  $-3$  are significantly shorter in duration than words in position  $-4$  (in all other languages, words in positions  $-3$  vs.  $-4$  are not significantly different in duration).

Estimated lengthening factors (Table 2) show that the degrees by which final words are lengthened vary greatly across languages, with between 7% (in Even) and 59% (in English) longer durations of final words compared to medial words. For English, this is in line with about 50% lengthening reported by Yuan et al. (2006).

Degrees of lengthening of penultimate words – where it occurs – are consistently much lower than those of final words, ranging from 2% in NlIng to 6% in Dutch, consistent with findings in the literature on English and Finnish (see Section 1). These values are in the same range as those of acceleration in the three languages where it occurs, with values of 2 to 4% for penultimate words and 2 to 3% for words in positions  $-3$  versus  $-4$ .

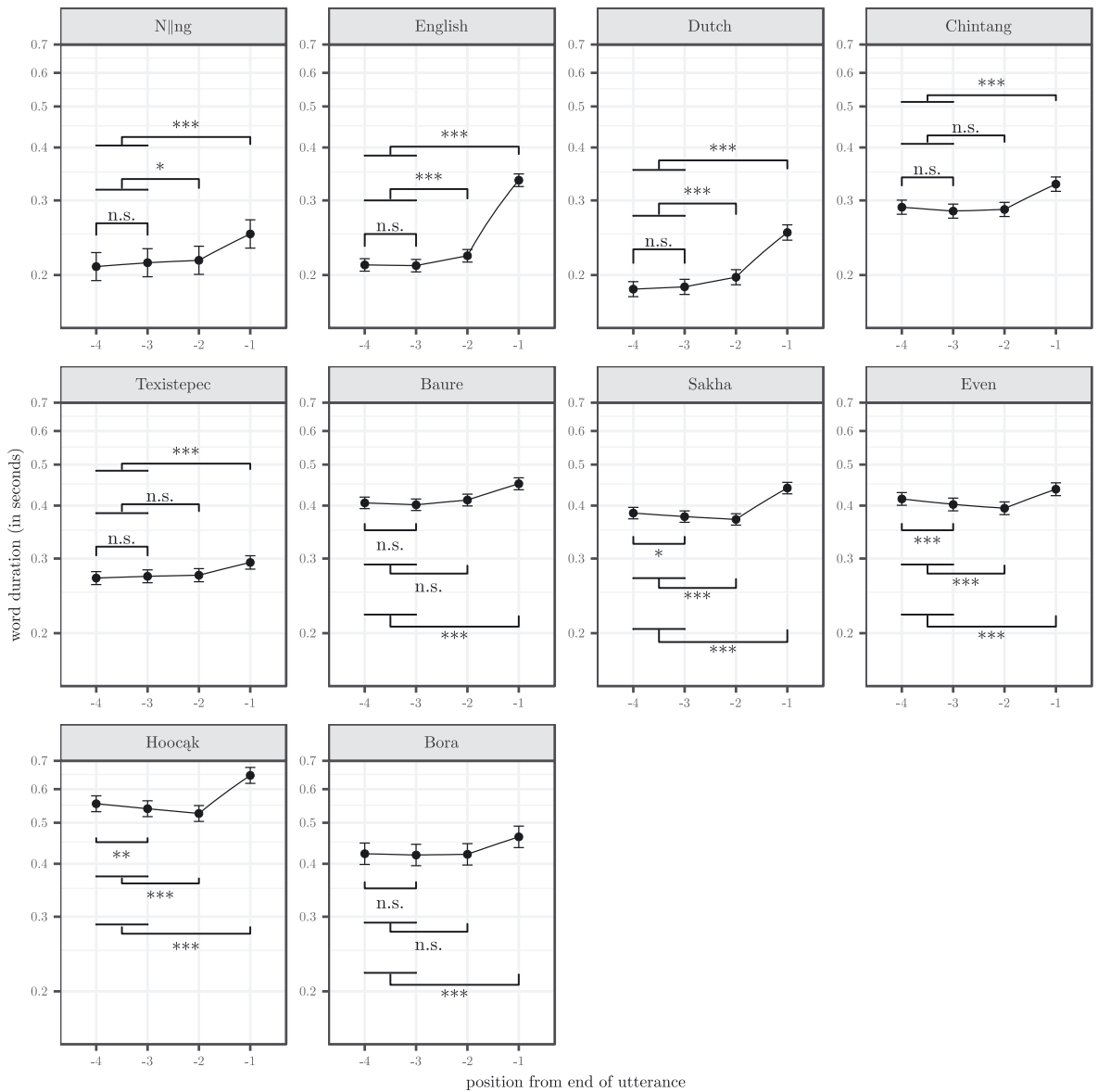
## 4 Discussion

Utterance-final words clearly exhibit lengthening effects across our sample. In three languages (English, Dutch, and NlIng) – but not in the others studied here – final lengthening appears to extend backwards across



word boundaries to penultimate words. For English, this is in line with earlier findings based on the same corpus we used (Yuan et al. 2006). This finding can be interpreted with respect to the average word length of the languages (see Table 1): English, Dutch, and Nlɪŋg have the shortest words, with on average between 3.45 (Nlɪŋg) and 3.85 segments (Dutch) per word, while words in other languages have on average at least 5.14 segments. Taking differences in expected duration as a function of average word length into account, we can generalize that across our 10 languages, final lengthening effects are detectable in words (−1 or −2) that start up to about 0.5 s before an utterance boundary.

Surprisingly, in three languages (Even, Sakha, and Hooc̣ək) there appears to be a pattern of acceleration up to and including the penultimate word within the utterance-final four-word windows studied here. It is noteworthy that these three languages have verb-final word order (Table 1). However, Bora, Chintang, Textistepec, and arguably Dutch, also have this word order without presenting acceleration. A more promising



**Figure 3:** Effect plots showing the predicted effect of word position on word duration based on the 10 multivariate models for the individual languages. Panels are ordered according to the average word length of the languages (cf. Table 1) from shortest to longest. Significance stars are based on the *t*-tests of the model coefficients calculated with a Satterthwaite approximation of the degrees of freedom using the R library *LmerTest* (cf. Table 2).

**Table 2:** Standardized  $\beta$  coefficients of the word position factor,  $p$ -values based on  $t$ -tests of the model coefficients calculated with a Satterthwaite approximation of the degrees of freedom using the R library `LmerTest`, and lengthening factors as the ratio of mean word durations of the positions being compared based on the predictions of the multivariate models.

	Difference among medial positions (-3 vs. -4)			Penultimate (-2) vs. medial (-3 and -4)			Final (-1) vs. medial (-3 and -4)		
	$\beta$	$p$	Factor	$\beta$	$p$	Factor	$\beta$	$p$	Factor
Baure	-0.007	0.571	0.99	0.010	0.390	1.02	0.069	0.000	1.12
Bora	-0.004	0.536	0.99	-0.002	0.777	1.00	0.059	0.000	1.10
Chintang	-0.015	0.174	0.98	-0.009	0.445	1.00	0.088	0.000	1.15
Dutch	0.008	0.269	1.01	0.043	0.000	1.06	0.204	0.000	1.35
English	-0.002	0.726	1.00	0.030	0.000	1.05	0.275	0.000	1.59
Even	-0.029	0.000	0.97	-0.049	0.000	0.96	0.051	0.000	1.07
Hoocak	-0.023	0.004	0.97	-0.045	0.000	0.96	0.130	0.000	1.18
Nllng	0.015	0.161	1.02	0.025	0.029	1.02	0.129	0.000	1.18
Sakha	-0.017	0.023	0.98	-0.030	0.000	0.98	0.118	0.000	1.16
Texistepec	0.007	0.537	1.01	0.012	0.321	1.01	0.065	0.000	1.08

explanation is again related to average word length, in combination with potential effects of (smaller) prosodic phrases such as intonational units within utterances: The three languages with apparent acceleration are the ones with the longest words in our sample, surpassed only by Bora, which displays no significant duration differences between words in position -4, -3, and -2. Having longer words means that it may be more likely that boundaries of (smaller) prosodic phrases occur within the four-word windows studied here. Our experience from qualitatively studying other languages with long words, such as Nunggubuyu (Northern Australia) and Athabaskan languages (North America) is that intonational phrases are rarely longer than two words in such languages. However, if boundaries of two-word prosodic phrases are not accompanied by a pause of at least 0.2 s, they are not detected by our method (see Section 2.1.2), but would have to be identified by other cues, such as pitch movements. Thus, some of the words in position -4 and -3 in our Even, Sakha, and Hoocak data may be final words of (intermediate) prosodic phrases, which would be followed by phrase-initial, faster words in positions -3 and -2, respectively, giving rise to apparent utterance-internal acceleration.

The cross-linguistic differences in degrees of utterance-final word lengthening can again be interpreted with respect to cross-linguistic differences of average word length in number of segments (Table 1). Since final segments within final words are expected to be disproportionately lengthened, the effect on the duration of final words as a whole, as measured here, should be magnified in languages with shorter words and thus with fewer segments. This explains the high degrees of final lengthening in English (on average 3.7 segments per word and 59% lengthening of final words) and Dutch (on average 3.85 segments per word and 35% lengthening of final words). However, words in Nllng are even shorter (on average 3.45 segments per word), while final words are only lengthened 18%, comparable to languages with long words like Hoocak, in which words are on average almost twice as long (6.64 segments). Additionally, Nllng, unlike other languages in our sample (see Table 1), does not have contrastive vowel length, which is known to reduce final lengthening effects (Nakai et al. 2009). This suggests that other factors are also involved in explaining degrees of final lengthening, which may include language-specific combinations of cues for signaling prosodic phrasing, text type (the English and Dutch corpora are different in that they are only conversational, while all other corpora are mostly narratives), and other differences in segment inventories or phonological and orthographic complexity.

## 5 Conclusion

This study provides cross-linguistic evidence for utterance-final lengthening effects through a comparison of 10 diverse languages: In all 10 languages studied here, utterance-final words have significantly longer

durations than utterance-medial words. While the universality of final lengthening has been widely assumed in the literature, based on reports from numerous languages and the lack of known counterexamples, to the best of our knowledge this has not been investigated previously in a single study comparing various languages using the same methods.

Both the extent and the degree of utterance-final word lengthening vary strongly between the languages studied here. Much of this variation can be explained by average word length: In general, if words are short on average, final words are more strongly affected and lengthening extends to penultimate words. This finding is in line with previous observations of disproportionately strong lengthening of final segments within final words. It also supports models that accommodate mechanisms by which final lengthening extends linearly backwards from utterance boundaries, also across word boundaries, possibly in addition to mechanisms that operate on specific structural positions, e.g. stressed syllables (Turk and Shattuck-Hufnagel 2007). Our results thus call for further cross-linguistic studies of lengthening within not only final, but also penultimate words, once phonetic, phonological, and syllable annotations become available and automatic phoneme alignment systems become more reliable for under resourced languages, too. Other cross-linguistic differences in our results, including the apparent acceleration up to the penultimate word of utterances in three out of the 10 languages studied here, call for the inclusion of other cues for the identification of prosodic phrases below the utterance level, such as pitch contours, in addition to pauses, in the study of utterance-final lengthening.

Finally, it is interesting to note that on all measures taken here, each of the three pairs of languages in our sample that are areally related appear to follow the same patterns (English and Dutch, Even and Sakha, and Bora and Baure – of which only English and Dutch are additionally genealogically related). This could suggest that patterns in utterance-final lengthening might be prone to areal spread, as pronunciation styles that traverse language boundaries in multilingual settings, although a much larger language sample would be needed to investigate this thoroughly.

As one step towards enabling such studies, the current paper demonstrates the potential of newly available audio and text materials from under-resourced languages for corpus phonetics (Lieberman 2019). Extracting more materials of this kind from language documentation collections held in archives like TLA<sup>1</sup> and ELAR<sup>2</sup> could substantially widen the cross-linguistic scope of corpus phonetics in terms of the genealogical, areal, and typological diversity of human languages.

**Acknowledgments:** The research of FS and JS was supported by a grant from the Volkswagen Foundation’s Dokumentation Bedrohter Sprachen (DoBeS) program (89 550). FS and BP are grateful to the LABEX ASLAN (ANR-10-LABX-0081) of Université de Lyon for its financial support within the program “Investissements d’Avenir” (ANR-11-IDEX-0007) of the French government operated by the National Research Agency (ANR). SW’s research was supported by JPICH/NWO, a subsidy of the Russian Government to support the Programme of Competitive Development of Kazan Federal University, and a major project from National Social Science Fund of China (no. 19ZDA300). We are grateful for helpful comments from Susanne Fuchs, Oksana Rasskazova, Colleen O’Brien, and two anonymous reviewers.

## References

- Aylett, Matthew & Alice Turk. 2004. The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech* 47(1). 31–56.
- Bates, Douglas, Martin Mächler, Benjamin M. Bolker & Steven C. Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48.
- Bell, Alan, Jason M. Brenier, Michelle Gregory, Cynthia Girand & Dan Jurafsky. 2009. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* 60(1). 92–111.

<sup>1</sup> <https://tla.mpi.nl>.

<sup>2</sup> <https://elar.soas.ac.uk>.

- Bell, Alan, Daniel Jurafsky, Eric Fosler-Lussier, Cynthia Girand, Michelle Gregory & Daniel Gildea. 2003. Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *The Journal of the Acoustical Society of America* 113(2). 1001–1024.
- Berkovits, Rochele. 1994. Durational effects in final lengthening, gapping, and contrastive stress. *Language and Speech* 37(3). 237–250.
- Bickel, Balthasar, Sabine Stoll, Martin Gaenszle, Novel Kishore Rai, Elena Lieven, Goma Banjade, Toya Nath Bhatta, Netra Paudyal, Judith Pettigrew, Ichchha P. Rai & Manoj Rai. 2011. *Audiovisual corpus of the Chintang language, including a longitudinal corpus of language acquisition by six children: Ca. 650,000 words transcribed and translated, of which ca. 450,000 glossed, plus paradigm sets and grammar sketches, ethnographic descriptions, photographs*. Nijmegen: The Language Archive. <https://hdl.handle.net/1839/00-0000-0000-0005-6F41-C@view> (accessed 20 March 2019).
- Brybaert, Marc & Boris New. 2009. Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods* 41(4). 977–990.
- Byrd, Dani, Jelena Krivokapić & Sungbok Lee. 2006. How far, how long: On the temporal scope of prosodic boundary effects. *Journal of the Acoustical Society of America* 120(3). 1589–1599.
- Byrd, Dani & Elliot Saltzman. 2003. The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics* 31(2). 149–180.
- Cambier-Langeveld, Tina. 1997. The domain of final lengthening in the production of Dutch. In Helen de Hoop & Jane A. Coerts (eds.), *Linguistics in the Netherlands 1997*, vol. 14, 13–24. Amsterdam: John Benjamins.
- CGN-consortium, Language and Speech Nijmegen & ELIS Gent. 2003. *Corpus Gesproken Nederlands*. Nijmegen: Nederlandse Taalunie.
- Cho, Taehong. 2016. Prosodic boundary strengthening in the phonetics–prosody interface. *Language and Linguistics Compass* 10(3). 120–141.
- Cho, Taehong, Jiseung Kim & Sahyang Kim. 2013. Preboundary lengthening and preaccentual shortening across syllables in a trisyllabic word in English. *Journal of the Acoustical Society of America* 133(5). EL384–EL390.
- Danielsen, Swintha, Franziska Riedel, Femmy Admiraal & Lena Terhart. 2009. *Baure documentation*. Nijmegen: The Language Archive. <https://hdl.handle.net/1839/00-0000-0000-000D-8382-B@view> (accessed 20 March 2019).
- Delattre, Pierre. 1966. A comparison of syllable length conditioning among languages. *International Review of Applied Linguistics* 7. 295–325.
- ELAN developers. 2019. *ELAN (Version 5.7) [Computer software] (June 14, 2019)*. Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. <https://tla.mpi.nl/tools/tla-tools/elan/> (accessed 20 July 2019).
- Fletcher, Janet. 2010. The prosody of speech: Timing and rhythm. In William J. Hardcastle, John Laver & Fiona E. Gibbon (eds.), *The handbook of phonetic sciences*, 2nd edn., 521–602. Chichester: Blackwell. <https://doi.org/10.1002/9781444317251.ch15>.
- Godfrey, John J., Edward C. Holliman & Jane McDaniel. 1992. SWITCHBOARD: Telephone speech corpus for research and development. *ICASSP-92: 1992 IEEE international conference on acoustics, speech, and signal processing*, vol. 1, 517–520.
- Güldemann, Tom, Martina Ernszt, Sven Siegmund & Alena Witzlack-Makarevich. 2011. *A text documentation of Nluu*. London: ELAR. <https://elar.soas.ac.uk/Collection/MPI194591> (accessed 20 March 2019).
- Hakokari, Jussi, Tuomo Saarni, Tapio Salakoski, Jouni Isoaho & Olli Aaltonen. 2005. Determining prepausal lengthening for Finnish rule-based speech synthesis. *Speech analysis, synthesis and recognition, applications of phonetics, September 19–23, 2005*. Kraków, Poland: AGH University of Science and Technology.
- Hammarström, Harald, Robert Forkel & Martin Haspelmath (eds.). 2018. *Glottolog 3.3*. Jena: Max Planck Institute for the Science of Human History. <https://doi.org/10.5281/zenodo.1321024>. <https://glottolog.org/> (accessed 20 March 2019).
- Hartmann, Iren. 2013. *Hoocqk corpus*. Leipzig: MPI-EVA.
- Himmelman, Nikolaus P. 1998. Documentary and descriptive linguistics. *Linguistics* 36(1). 161–195.
- Himmelman, Nikolaus P. & D. Robert Ladd. 2008. Prosodic description: An introduction for fieldworkers. *Language Documentation & Conservation* 2(2). 244–274. <http://hdl.handle.net/10125/4345> (accessed 20 March 2019).
- Himmelman, Nikolaus P., Meytal Sandler, Jan Strunk & Volker Unterladstetter. 2018. On the universality of intonational phrases: A cross-linguistic interrater study. *Phonology* 35(2). 207–245.
- Hyman, Larry M. 2013. Penultimate lengthening in Bantu. In Balthasar Bickel, Lenore A. Grenoble, David A. Peterson & Alan Timberlake (eds.), *Language typology and historical contingency. In honor of Johanna Nichols* (Typological Studies in Language 104), 309–330. Amsterdam: John Benjamins.
- Jun, Sun-Ah. 2014. Prosodic typology: By prominence type, word prosody, and macro-rhythm. In Sun-Ah Jun (ed.), *Prosodic typology II: The phonology of intonation and phrasing*, 520–540. Oxford: Oxford University Press.
- Keating, Patricia, Taehong Cho, Cecile Fougeron & Chai-Shune Hsu. 2003. Domain-initial articulatory strengthening in four languages. In John Local, Richard Ogden & Rosalind Temple (eds.), *Phonetic interpretation: Papers in laboratory phonology VI*, 143–161. Cambridge: Cambridge University Press.
- Kisler, Thomas, Florian Schiel & Han Sloetjes. 2012. Signal processing via web services: The use case WebMAUS. In *Proceedings of digital humanities 2012*, 30–34. Hamburg.
- Lieberman, Mark Y. 2019. Corpus phonetics. *Annual Review of Linguistics* 5(1). 91–107.

- Lindblom, Björn. 1968. Temporal organization of syllable production. *Speech transmission lab. Quarterly progress status report* 2. 1–5.
- Lindblom, Björn. 1979. Final lengthening in speech and music. In Eva Gårding, Gösta Bruce & Robert Bannert (eds.), *Nordic prosody (Travaux de l'Institut de Linguistique de Lund 13)*, 85–101. Lund: Lund University.
- Michelas, Amandine & Mariapaola D'Imperio. 2010. Durational cues and prosodic phrasing in French: Evidence for the intermediate phrase. In *Proceedings of speech prosody 2010*, paper 881. Chicago, USA. <https://hal.archives-ouvertes.fr/hal-00463205> (accessed 20 March 2019).
- Nakai, Satsuki, Sari Kunnari, Alice Turk, Kari Suomi & Riikka Ylitalo. 2009. Utterance-final lengthening and quantity in Northern Finnish. *Journal of Phonetics* 37(1). 29–45.
- Oller, D. Kimbrough. 1979. Syllable timing in Spanish, English, and Finnish. In Harry Francis Hollien & Patricia Hollien (eds.), *Current issues in the phonetic sciences: Proceedings of the IPS-77 congress, Miami Beach, Florida, 17–19th December 1977 (Amsterdam studies in the theory and history of linguistic science. Series IV, Current issues in linguistic theory 9)*, 331–341. Amsterdam: John Benjamins. <https://doi.org/10.1075/cilt.9.38oll>.
- Ordin, Mikhail, Leona Polyanskaya, Itziar Laka & Marina Nespors. 2017. Cross-linguistic differences in the use of durational cues for the segmentation of a novel language. *Memory & Cognition* 45(5). 863–876.
- Pakendorf, Brigitte (ed.). 2007. *Documentation of Sakha (Yakut)*. Leipzig: MPI-EVA.
- Pakendorf, Brigitte, Dejan Matić, Natalia Aralova & Alexandra Lavrillier. 2010. *Documentation of the dialectal and cultural diversity among Évens in Siberia*. Nijmegen, Leipzig: DOBES, MPIP, MPI-EVA.
- Piantadosi, Steven T., Harry Tily & Edward Gibson. 2011. Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences of the United States of America* 108(9). 3526–3529.
- Plag, Ingo, Julia Homann & Gero Kunter. 2017. Homophony and morphology: The acoustics of word-final S in English. *Journal of Linguistics* 53(1). 181–216.
- R Core Team. 2018. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. <http://www.R-project.org> (accessed 20 March 2019).
- Šavrič, Bojan, Tom Patterson & Bernhard Jenny. 2019. The equal earth map projection. *International Journal of Geographical Information Science* 33(3). 454–465.
- Seifart, Frank. 2009. Bora documentation. In Frank Seifart, Doris Fagua, Jürg Gasché & Juan Alvaro Echeverri (eds.), *A multimedia documentation of the languages of the People of the Center. Online publication of transcribed and translated Bora, Ocaina, Nonuya, Resígaro, and Witoto audio and video recordings with linguistic and ethnographic annotations and descriptions*. Nijmegen: The Language Archive. <https://hdl.handle.net/1839/00-0000-0000-0008-38E5-2> (accessed 20 March 2019).
- Seifart, Frank, Jan Strunk, Swintha Danielsen, Iren Hartmann, Brigitte Pakendorf, Søren Wichmann, Alena Witzlack-Makarevich, Nivja H. de Jong & Balthasar Bickel. 2018. Nouns slow down speech across structurally and culturally diverse languages. *Proceedings of the National Academy of Sciences of the United States of America* 115(22). 5720–5725.
- Seyfarth, Scott. 2014. Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition* 133(1). 140–155.
- Shattuck-Hufnagel, Stefanie & Alice E. Turk. 1996. A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research* 25(2). 193–247.
- Sóskuthy, Márton & Jennifer Hay. 2017. Changing word usage predicts changing word durations in New Zealand English. *Cognition* 166. 298–313.
- Strunk, Jan, Florian Schiel & Frank Seifart. 2014. Untrained forced alignment of transcriptions and audio for language documentation corpora using WebMAUS. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the ninth international conference on language resources and evaluation (LREC 2014)*, 3940–3947. Reykjavik: European Language Resources Association (ELRA). [http://www.lrec-conf.org/proceedings/lrec2014/pdf/1176\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2014/pdf/1176_Paper.pdf) (accessed 20 March 2019).
- Strunk, Jan, Frank Seifart, Swintha Danielsen, Iren Hartmann, Brigitte Pakendorf, Søren Wichmann, Alena Witzlack-Makarevich & Balthasar Bickel. 2020. Determinants of phonetic word duration in ten language documentation corpora: Word frequency, complexity, position, and part of speech. *Language Documentation & Conservation* 14. 423–461. <http://hdl.handle.net/10125/24926> (accessed 20 July 2020).
- Tang, Kevin & Ryan Bennett. 2018. Contextual predictability influences word and morpheme duration in a morphologically complex language (Kaqchikel Mayan). *Journal of the Acoustical Society of America* 144(2). 997–1017.
- Turk, Alice E. & Stefanie Shattuck-Hufnagel. 2000. Word-boundary-related duration patterns in English. *Journal of Phonetics* 28(4). 397–440.
- Turk, Alice E. & Stefanie Shattuck-Hufnagel. 2007. Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics* 35(4). 445–472.
- Turk, Alice E. & Laurence White. 1999. Structural influences on accentual lengthening in English. *Journal of Phonetics* 27(2). 171–206.
- Vaissière, Jacqueline. 1983. Language-independent prosodic features. In Anne Cutler & D. Robert Ladd (eds.), *Prosody: Models and measurements (Springer Series in Language and Communication 14)*, 53–66. Springer: Heidelberg. [https://doi.org/10.1007/978-3-642-69103-4\\_5](https://doi.org/10.1007/978-3-642-69103-4_5).

- Wichmann, Søren. 1996. *Cuentos y colorados en popoluca de Texistepec*. Copenhagen: C.A. Reitzel.
- Wightman, Colin W., Stefanie Shattuck-Hufnagel, Mari Ostendorf & Patti J. Price. 1992. Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America* 91(3). 1707–1717.
- Yuan, Jiahong, Mark Liberman & Christopher Cieri. 2006. Towards an integrated understanding of speaking rate in conversation. *Interspeech 2006*, 541–544. [https://www.isca-speech.org/archive/archive\\_papers/interspeech\\_2006/i06\\_1795.pdf](https://www.isca-speech.org/archive/archive_papers/interspeech_2006/i06_1795.pdf) (accessed 20 March 2019).

---

**Supplementary material:** The online version of this article offers supplementary material (<https://doi.org/10.1515/lingvan-2019-0063>).